UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/033,772	12/28/2001	James F. Arnold	SRI/4565-1	9261

52197 7590 05/24/2005

MOSER, PATTERSON & SHERIDAN, LLP
SRI INTERNATIONAL
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702

EXAMINER	
LERNER, MARTIN	

ART UNIT	PAPER NUMBER
2654	

DATE MAILED: 05/24/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/033,772	ARNOLD ET AL.	
	Examiner	Art Unit	
	Martin Lerner	2654	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *15 February 2005*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1 to 31* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1 to 31* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

Claim Rejections - 35 USC § 102

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (a) the invention was known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the invention thereof by the applicant for a patent.

2. Claims 1, 5 to 7, 9, 13 to 16, 20, 22, and 26 to 31 are rejected under 35 U.S.C. 102(a) as being anticipated by *Thrift et al.*

Regarding independent claims 1, 15, 16, and 30, *Thrift et al.* discloses a method, system, and computer-readable medium, comprising:

"receiving a speech signal locally from a user via a client device" – microphone 10b receives voice input from a user; voice activated control unit 10 ("a client device") has microphone 10b (column 2, lines 59 to 62: Figure 1);

"performing speech recognition on said speech signal in accordance with an embedded speech recognizer of said client device to produce a recognizable text signal, wherein said embedded speech recognizer employs a language model" – in one embodiment, control unit 10 performs all of the voice recognition process and delivers speech data to host computer 11 via transmitter 10g (column 3, lines 1 to 3: Figure 1); if control unit 10 performs all voice recognition processes, memory 10f stores these processes (as a voice recognizer) as well as grammar files (column 3, lines 22 to 45: Figure 1); broadly, grammar files are "a language model"; implicitly, speech data is in

the form of "a recognizable text signal" because speech recognition generates text from speech;

"adapting said performance of speech recognition based on at least one local parameter of said speech signal" – memory 10f stores a grammar file generator for dynamically generating a grammar (column 3, lines 41 to 45: Figure 1); grammars for speakable links may be dynamically created so that only the grammar for a current display is active and is updated when a current display is generated; dynamic grammar creation reduces the amount of required memory 10f; dynamic grammar files are created from current Web pages; every time the screen 40 changes, the user agent 64 creates a grammar containing the currently visible links (column 5, line 48 to column 6, line 25: Figure 5); dynamic updating of grammar files every time a screen changes is equivalent to "adapting said performance of speech recognition", where changing of a screen is based on "at least one local parameter of said speech signal";

"forwarding said recognizable text signal to a remote server" – the output of the voice recognizer is speech data; the speech data is transmitted to host system 11 ("a remote server"), which performs voice control interpretation processes (column 3, lines 45 to 56: Figure 1).

Regarding independent claims 9, 22, and 31, *Thrift et al.* discloses a method, server, and computer-readable medium, comprising:

"receiving a recognizable text signal representative of a user speech signal from a client device, wherein said recognizable text is generated using a speech recognizer

having a language model on said client device" – microphone 10b receives voice input from a user; voice activated control unit 10 ("a client device") has microphone 10b (column 2, lines 59 to 62: Figure 1); in one embodiment, control unit 10 performs all of the voice recognition process and delivers speech data to host computer 11 via transmitter 10g (column 3, lines 1 to 3: Figure 1); if control unit 10 performs all voice recognition processes, memory 10f stores these processes (as a voice recognizer) as well as grammar files (column 3, lines 22 to 45: Figure 1); broadly, grammar files are "a language model"; implicitly, speech data is in the form of "a recognizable text signal" because speech recognition generates text from speech;

"wherein said recognizable text is generated in accordance with adapting said performance of speech recognition based on at least one local parameter of said speech signal" – memory 10f stores a grammar file generator for dynamically generating a grammar (column 3, lines 41 to 45: Figure 1); grammars for speakable links may be dynamically created so that only the grammar for a current display is active and is updated when a current display is generated; dynamic grammar creation reduces the amount of required memory 10f; dynamic grammar files are created from current Web pages; every time the screen 40 changes, the user agent 64 creates a grammar containing the currently visible links (column 5, line 48 to column 6, line 25: Figure 5); dynamic updating of grammar files every time a screen changes is equivalent to "adapting said performance of speech recognition", where changing of a screen is "at least one local parameter";

"processing said recognizable text signal in accordance with a task model" – the output of the voice recognizer is speech data; the speech data is transmitted to host system 11 ("a remote server"), which performs voice control interpretation processes; examples of voice control interpretation are web browsing and commands to a television (column 3, lines 45 to 65: Figure 1); web browsing and commands to a television are examples of "a task model".

Regarding claims 5, 13, 20, and 26, *Thrift et al.* discloses host 11 ("said remote server") could dynamically generate the grammar and download the grammar file to control unit 10 (column 3, lines 41 to 45: Figure 1); a grammar file is downloaded in response to speech data ("said recognizable text signal") requesting a new web page (column 5, line 48 to column 6, line 13: Figure 5).

Regarding claim 6, *Thrift et al.* discloses the output of the voice recognizer is speech data; the speech data is transmitted to host system 11 ("a remote server"), which performs voice control interpretation processes; examples of voice control interpretation processes are web browsing and commands to a television (column 3, lines 45 to 65: Figure 1); web browsing and commands to a television are examples of "a task model".

Regarding claims 7, 14, and 28, *Thrift et al.* discloses examples of voice control interpretation are web browsing and commands to a television; host system 11 ("a remote server") may respond to voice input to control unit 10 by executing a command or providing a hypermedia (Web) link (column 3, lines 45 to 65: Figure 1); thus, host

system 11 must monitor "progress toward satisfying a goal of said user" to display a television schedule or browse the web.

Regarding claim 27, *Thrift et al.* discloses host 11 ("said remote server") could dynamically generate the grammar and download the grammar file to control unit 10 (column 3, lines 41 to 45: Figure 1); a grammar file is downloaded is response to speech data ("said recognizable text signal") requesting a new web page (column 5, line 48 to column 6, line 13: Figure 5); implicitly, something that forwards grammar file updates from a host system 11 to a control unit 10 is "a grammar manager".

Regarding claim 29, *Thrift et al.* discloses host system 11 provides voice control interpretation processes for dialogs via speakable hotlist processes (column 4, line 33 to column 5, line 19: Figure 3); an interpretation process for determining which processes are hotlist processes is equivalent to "a dialog manager".

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

4. Claims 2 to 4, 10 to 12, 17 to 19, and 23 to 25 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Thrift et al.* in view of *Balakrishnan et al.*

Thrift et al. updates a grammar file ("adapting said performance of speech recognition") based upon a currently displayed web page of a speakable command list

("based on at least one local parameter"), but omits adapting performance of speech recognition based on a parameter representative of environmental noise, acoustic environment, and pronunciation of a user. However, it is well known that speech recognition systems can be trained to improve performance with respect to individual user pronunciations and environmental noise. *Balakrishnan et al.* teaches context dependent phoneme networks that are specific to a user and an environment. (Column 2, Lines 10 to 49) In operation, a first part of an operating system 44 generates a CD phoneme network in order to capture user and environment specific acoustic models, which are continually adapting to the user's voice, environment, and use of language. The second part 50 of the operating system 44 then uses appropriate search engine applets 51 to retrieve a CD network. (Column 4, Line 66 to Column 5, Line 56) Implicitly, an environment for speech recognition is inclusive of environmental noise. The objective is to eliminate obstacles to computer speech recognition by not requiring that each application will have to keep separate acoustic models for each user/environment and so that performance is not sacrificed. (Column 1, Lines 24 to 55) It would have been obvious to one having ordinary skill in the art to adapt performance of speech recognition based on parameters representative of environmental noise, acoustic environment, and pronunciation of a user as taught by *Balakrishnan et al.* in the wireless voice-activated device for control of a processor-based host system of *Thrift et al.* for the purpose of eliminating obstacles to speech recognition by not requiring that each application have separate acoustic models for each user/environment.

5. Claims 8 and 21 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Thrift et al.* in view of *Ramaswamy et al.*

Thrift et al. discloses grammar files ("said language model") are stored in memory 10f of control unit 10 ("said client device"), but does not specifically say that grammar files are stored in a cache. However, it is well known that files currently being used by a computer system are commonly stored in cache to reduce memory access operations. Thus, it is likely implicit that memory 10f includes a cache, and grammar files are stored in cache memory for *Thrift et al.* *Ramaswamy et al.* teaches an analogous art speaker verification method and system, where speech recognition engines use a language model. When more than one language model is used, some of the models may be personalized to a given user, and stored in a personal cache, built using words and phrases spoken frequently by a given user. (Column 5, Lines 22 to 27) It would have been obvious to one having ordinary skill in the art to store dynamically updated grammar files of control unit 10 from *Thrift et al.* in a cache memory as suggested by *Ramaswamy et al.* for the purpose of reducing memory access operations for words and phrases spoken frequently by a given user.

Response to Arguments

6. Applicants' arguments filed 15 February 2005 have been fully considered but they are not persuasive.

Firstly, Applicants argue that *Thrift et al.* does not teach that the grammar used by the remote control for speech recognition is dynamically updated based on a local parameter of the input user command (*e.g.* speech signal). This position is traversed.

Thrift et al. teaches dynamically updating the grammar based upon a parameter of the speech signal because the user agent creates a grammar in response to a user speaking an underlined link. A user of a voice control unit 10 may speak a link from a page being displayed on display 10a. (Column 5, Lines 18 to 19) Speech module 66 inputs the user's speech into user agent 64. (Column 6, Lines 2 to 3) Then, every time screen 40 changes, the user agent creates a grammar containing the currently visible underlined phrases (links). (Column 6, Lines 8 to 10) The grammars for speakable links are dynamically created so that only the grammar for a current display is active and updated when a new current display is generated. Dynamic grammar creation reduces the amount of required memory 10f. (Column 5, Lines 48 to 52) Changing the grammar by a user agent in response to a spoken link is equivalent to "adapting performance of speech recognition". Changing a grammar in response to a spoken link is also "based on at least one local parameter" because a user agent must supply "a local parameter" to voice activated control unit 10 in order to cause a grammar to be updated for a new display. The local parameter is "of said speech signal" because a direction to change a grammar is in response to a spoken link.

Secondly, Applicants argue that *Thrift et al.* only teaches a speech recognition grammar for remote control of a host that is dynamically generated based on a current display (*e.g.* of Web pages or links) on the host. Applicants state that the dynamically

created grammar of *Thrift et al.* is based on what the user might say (*e.g.* valid commands based on the current display), and not what the user has already said in the an input speech signal. Moreover, Applicants maintain that they positively claim adapting the speech recognition process based on at least one local parameter of the input speech signal (*e.g.* environmental noise, user pronunciation and the like). This position is traversed.

Applicants have not expressly claimed that the local parameter of the input speech signal is limited to environmental noise or user pronunciation. Applicants are unwarrantedly attempting to read limitations into the claims from their Specification. Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993). Here, the claims do not expressly state that the local parameter of the speech signal is limited to environmental noise or user pronunciation. Thus, the phrase "at least one local parameter of the speech signal" should be broadly construed. During patent examination, the pending claims must be "given their broadest reasonable interpretation consistent with the specification." *In re Hyatt*, 211 F.3d 1367, 1372, 54 USPQ2d 1664, 1667 (Fed. Cir. 2000). Applicant always has the opportunity to amend the claims during prosecution, and broad interpretation by the examiner reduces the possibility that the claim, once issued, will be interpreted more broadly than is justified. *In re Prater*, 415 F.2d 1393, 1404-05, 162 USPQ 541, 550-51 (CCPA 1969)

Moreover, Applicants' attempt to distinguish adapting performance of speech recognition based upon what the user has already said instead of what the user might say is not persuasive. In *Thrift et al.*, what the user has already said determines what page is currently displayed, and thus, which grammar is active. The active grammar is based on what a user has already said in *Thrift et al.*, because what a user has already said determines which web page is active, and which grammar is currently active.

Thirdly, Applicants argue that *Balakrishnan et al.* teaches away from a combination with *Thrift et al.* because *Thrift et al.* teaches dynamically adapting a language model based on current information, and *Balakrishnan et al.* teaches dynamically adapting an acoustic model or phoneme network based on current information for use with a static language model. Thus, Applicants submit that the rejection is based upon hindsight. This position is traversed.

Dynamic adaptation of a static language model is equivalent to dynamic adaptation of a language model. Applicants' attempt to distinguish a dynamic adaptation of a static language model from a dynamic adaptation of a language model is mere semantics. A language model is not static if it is being adapted. In fact, *Balakrishnan et al.* discloses models that are "continuously adapting to the user's voice, environment and use of language." (Column 5, Lines 5 to 6) Furthermore, the combination of *Thrift et al.* and *Balakrishnan et al.* is not based upon hindsight and does not teach away. *Thrift et al.* and *Balakrishnan et al.* are both from the same field of endeavor relating to speech recognition in a network environment. A proper motivation is provided for *prima facie* obviousness, as *Balakrishnan et al.* teaches an objective to

eliminate the need for storing separate acoustic models for each user/environment. (Column 1, Lines 24 to 55)

Fourthly, Applicants argue that *Ramaswamy et al.* teaches away from a combination with *Thrift et al.*, as *Ramaswamy et al.* discloses a language model that is dependent on past data (*e.g.* stored user behavior patterns), whereas Applicants provide a comparison against a current speaker's behavior. This is not persuasive.

Ramaswamy et al. is merely cited for well-known feature of a cache in speech recognition. Indeed, caches are commonly used in many computer systems, but *Ramaswamy et al.* teaches that caches are used for personalized language models for frequently spoken words and phrases. *Thrift et al.* and *Ramaswamy et al.* are both from same field of endeavor, *i.e.* speech recognition. Applicants are attacking the references individually without consideration of the reasons for the combination. One cannot show nonobviousness by attacking references individually where the rejections are based on combinations of references. See *In re Keller*, 642 F.2d 413, 208 USPQ 871 (CCPA 1981); *In re Merck & Co.*, 800 F.2d 1091, 231 USPQ 375 (Fed. Cir. 1986).

Therefore, the rejections of claims 1, 5 to 7, 9, 13 to 16, 20, 22, and 26 to 31 under 35 U.S.C. 102(a) as being anticipated by *Thrift et al.*, of claims 2 to 4, 10 to 12, 17 to 19, and 23 to 25 under 35 U.S.C. 103(a) as being unpatentable over *Thrift et al.* in view of *Balakrishnan et al.*, and of claims 8 and 21 under 35 U.S.C. 103(a) as being unpatentable over *Thrift et al.* in view of *Ramaswamy et al.*, are proper.

Conclusion

7. **THIS ACTION IS MADE FINAL.** Applicants are reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Martin Lerner whose telephone number is (703) 308-9064. The examiner can normally be reached on 8:30 AM to 6:00 PM Monday to Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (703) 305-9645. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

ML
5/12/05

Martin Lerner
Examiner
Group Art Unit 2654